Assessment Schedule – 2010

Scholarship Statistics and Modelling (93201)

Evidence Statement

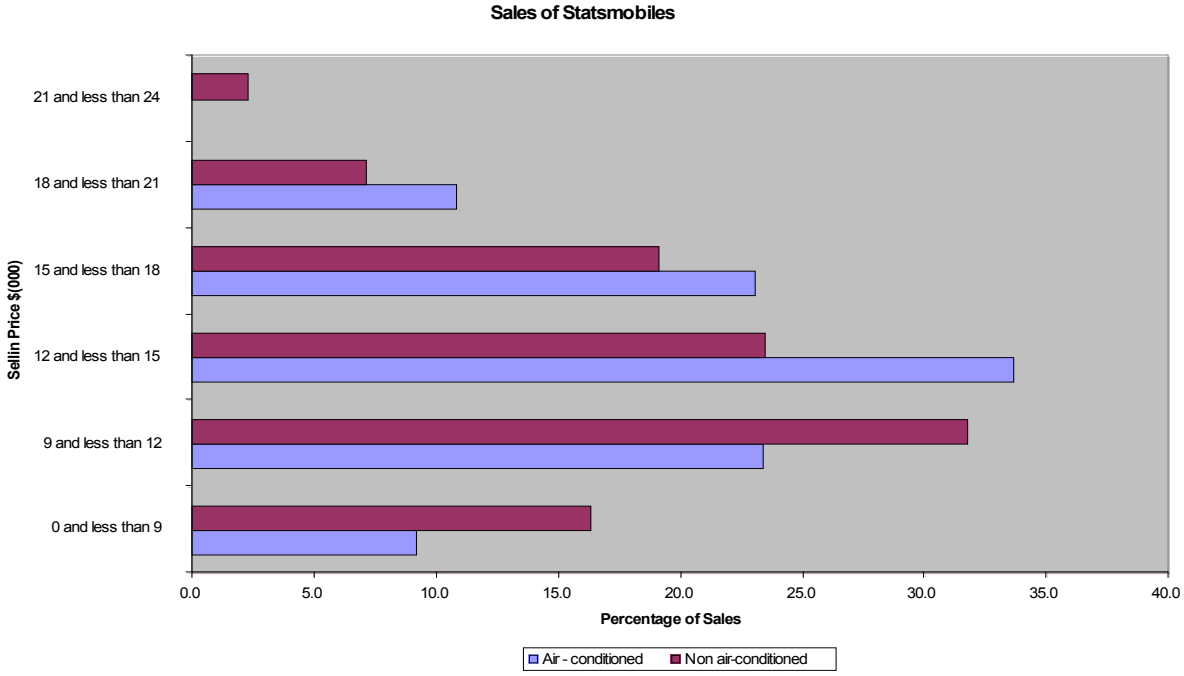
General Principles:

1. Ignore incorrect answers if alongside correct answers. The exception is contradictory statements.
2. Ignore minor copying errors.

QUESTION ONE

(a)

Evidence:
A multiple bar chart provides a comparison of the percentage of sales between air-conditioned and non air-conditioned second-hand statsmobiles.



Distribution of selling prices:
Air-conditioned cars – roughly symmetrical
Non air-conditioned cars – Skewed to the right

Ranges:
Air-conditioned: Maximum $21 000
Non Air-conditioned: Maximum $24 000

Central Measures:
Median: $1 305 higher for air-conditioned statsmobiles
Mean: $445 higher for air – conditioned statsmobiles

Spread:
The standard deviation is $310 higher for non air-conditioned statsmobiles so the non air-conditioned statsmobile prices are more spread out about the mean.

Modal Class:
$12 000 and less than $15 000 for air-conditioned statsmobiles
$9 000 and less than $12 000 for non air-conditioned statsmobiles

Note:

1. All statements have to be comparative and distinct.
2. Can have other comparative statements.
3. A general comparative statement must be backed up by descriptive statistics.
4. Graph alternatives could be either two side-by-side box plots or two percentage cumulative graphs (ogives) on the same set of axes.
5. Graph must distinguish between air-conditioned and non air-conditioned statsmobiles.

Judgement:

S: Graph correct plus 3 comparative statements.
P: Graph correct only or fewer than 3 comparative statements.

(b)(i)

Evidence:
Using the sales records stratify according to air-conditioned or non air-conditioned statsmobiles. Select 200 records at random from the air-conditioned statsmobiles and 300 records at random from the non air-conditioned statsmobiles. Allocate each statsmobile a random number according to its listing on the data base. Use these numbers to make random selections from each stratum ignoring repetitions and numbers outside range.

Note:

1. Random sampling on all the sales records isn't acceptable.
2. First two sentences in evidence not complete enough.
3. Systematic sampling within each type of statsmobile with appropriate explanation of random selection within starting value is acceptable.

(b)(ii)

Evidence:
The 95% confidence interval for the difference in the population mean selling prices is given by:

$$\mu_{ac} - \mu_{nac} = (15513 - 12581) \pm 1.96\sqrt{\frac{3896^2}{200} + \frac{4074^2}{300}}$$

So \$2 222 < $\mu_{ac} - \mu_{nac}$ < \$3 642 or $\mu_{ac} - \mu_{nac}$ = \$2 932 ± \$710

Note:

1. Rounding errors accepted.
2. Can have $\mu_{nac} - \mu_{ac}$ = –\$2 932 ± \$710

It's likely that the difference in the population means between the prices of air-conditioned statsmobiles to non air-conditioned statsmobiles is within the interval \$2 222 and \$3 642.
As zero isn't contained within this confidence interval we conclude that there is a statistically significant difference in the mean second hand prices between air-conditioned and non air-conditioned statsmobiles.

Judgement:

S: (i) or (ii) correct.
P: (ii) calculation only

(b)(iii)

Evidence:

Set $1.96\sqrt{\frac{3896^2}{200+2p}+\frac{4074^2}{300+3p}} = 558$

81 050.60 (100 + p) = 13 121 900

p = 61.90% so sample sizes need to be increased by 62%.

Judgement:

O: Correct answer
P: Some evidence of correct method.

QUESTION TWO

(a)

Evidence:

Pr (Total demand over 6 months $\geq$ 3360) = pr (average monthly demand $\geq$ 3360/6)
Pr $(z \geq \frac{560-600}{72/\sqrt{6}})$ = pr ($z \geq -1.361$) = 0.5 + 0.4133 = 0.9133.
Assumption: The monthly demands are independent of each other.

Note:

1. Can use totals to evaluate probability with mean 3600 and standard deviation $\sigma_T = \sqrt{6 \times 72^2} = 72\sqrt{6}$

Judgement

O: Both probability and assumption correct.
P: Probability correct only.

(b)

Evidence:
Let x = number of model X manufactured annually
Let y = number of model Y manufactured annually
The constraints are:

- $x + y \geq 6000$
- $1.7x \geq y \geq 0.95x$ or $y \leq 1.7x$ and $y \geq 0.95x$

The cost function has the form: $C = ay + x$ where for each case:

(i) $a < 1$
(ii) $a = 1$
(iii) $a > 1$

To ensure minimum cost C in each case, the optimal solutions are respectively:

(i) x = 2 223 and y = 3 777
(ii) Integer points (x,y) where 2 223 $\leq x \leq$ 3 777 and $x + y$ = 6 000
(iii) x = 3 076 and y = 2 924

Judgement

S: At least two optimal solutions correct.
P: Constraints and cost function correctly stated and/or one optimal solution correct.

(c)

Evidence:
Let x = number of statsmobiles requiring fixing
Pr ($x \leq 1$) = 0.2
So using Binomial with p = 0.03,
$0.97^{n} + n\ 0.97^{n-1}\ 0.03 = 0.2$
$n = 99$

Or using Poisson approximation to the Binomial with $\lambda = np = 0.03n$,
for pr ($x \leq 1$) = 0.2 so $\lambda = 3$ giving n = 100

Judgement

S: Correct value for n
P: Some evidence for correct choice of probability distribution with associated method.

QUESTION THREE

(a)

Evidence:
Let p = required proportion to be found.
Pr (Statsmobile has at least three doors given it's air-conditioned)
= Pr (Statsmobile is air-conditioned and has at least three doors) / Pr (Statsmobile is air-conditioned)

$$p = \frac{0.3}{0.45(1-p)+0.3}$$

$3p^2 - 5p + 2 = 0$ so either $p = 2/3$ or $p = 1$.

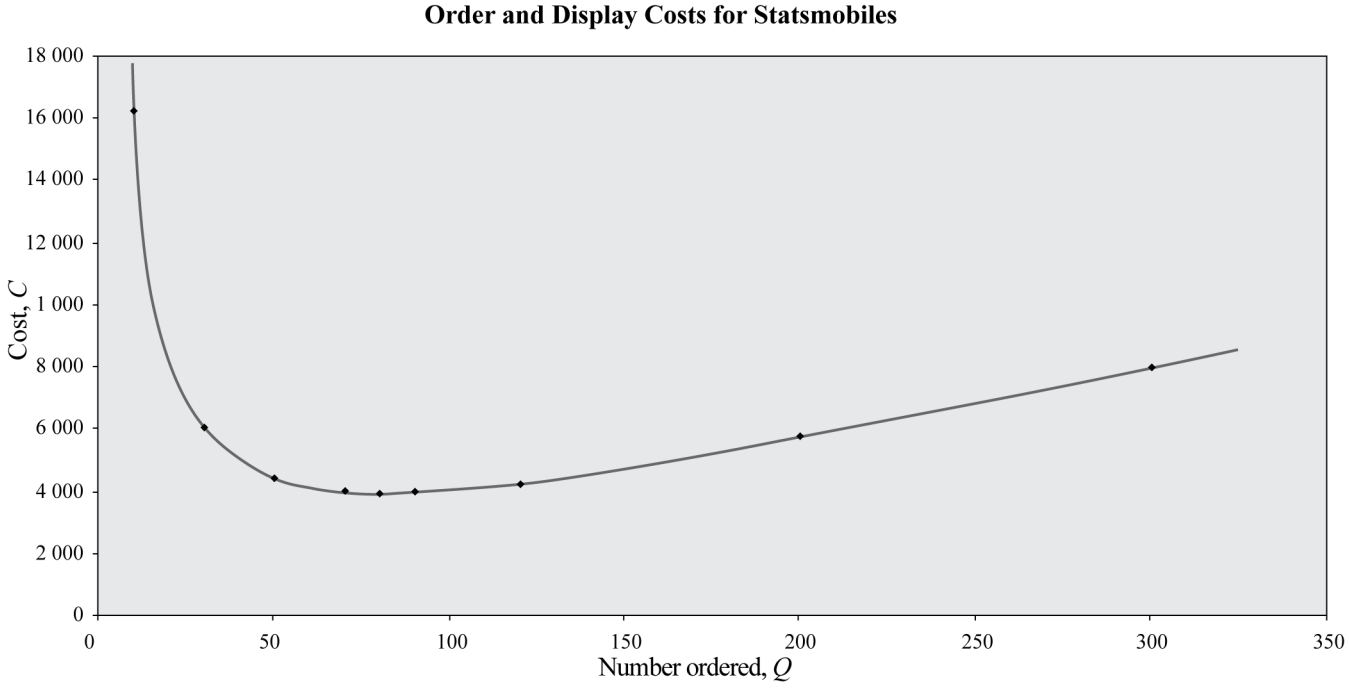
Now reject p = 1 solution as we have some two-door statsmobiles.

Judgement

O: Correct answer with relevant working.
P: Some indication of correct method.

(b)(i)

Evidence:



Use six pairs of simultaneous equations to solve for a and b.
AB: $a = 0.0258$ and $b = 156.916$
BC: $a = 0.0248$ and $b = 161.96$
AC: $a = 0.025$ and $b = 158.67$
AD: $a = 0.0249$ and $b = 159.0$
BD: $a = 0.0248$ and $b = 161.93$
CD: $a = 0.0248$ and $b = 161.87$
Taking the averages we get a =0.025 to 3 s.f. and b = 160 to 3 s.f.

Note:

1. As Q tends to zero we have a vertical asymptote.
2. As Q tends to infinity we have $C = a\,Q$ as an asymptote.
3. We can transform equation to $Y = b + aX$ where $Y = CQ$ and $X = Q^2$. The value of a is obtained by the gradient of the resulting line fit and the value of b from the Y intercept.
4. Partially correct if either the graph is correct or both a and b are correct.

(b)(ii)

Evidence:
Use graphics calculator directly to get $Q = 80$ and C min = \$4 000.

Note:

1. Can differentiate and equate $\frac{dC}{dQ}$ to zero to get Q = 80. Confirm minimum by $\frac{d^2C}{dQ^2} > 0$ when $Q = 80$. Then find C minimum by substitution.
2. Partially correct if some indication of correct method is shown.

Judgement

2S: Both (i) and (ii) correct.
S +P: Either (i) or (ii) completely correct with the other partially correct.
S: Either (i) or (ii) correct.
2P: Both (i) and (ii) partially correct.
P: Either (i) or (ii) partially correct.

QUESTION FOUR

(a)

Evidence:

1. The relationship between price and age is negative.
2. The relationship between price and age is moderately non-linear with minimal scatter.
3. The ages range from one to 14 years with a decreasing price range of \$24K to \$5K.
4. The relationship between price and odometer reading is negative.
5. The relationship between price and odometer reading is weak non-linear with a lot of scatter.
6. The odometer readings vary from \$38K to \$184K over a decreasing price range of \$24K to \$5K.
7. The relationship between price and odometer reading isn't as linear as the relationship between price and age.

Note:
Non comparative statements okay.

Judgement

S: Four or more points correct.
P: Three to one points correct**.**

(b)

Evidence:
Use the price versus age graph as a lot less scatter is evident compared to the price versus odometer graph. With the price versus age graph, the curve appears to be the best choice in terms of fit so:
Predicted Price = $25.4 \times e^{-0.13 \times 7} = \$10\ 224$

The validity is reasonable as the scatter is low with correlation moderately negative and curve fit reasonable. Using the line would give a prediction of $11 700 which seems too high by observation.

Judgement

S: Reasonable prediction and validity comment correct.
P: Reasonable prediction only.

(c)

Evidence:
Other factors that could be considered are:

1 CC Rating (Engine Size)
2. Reputation of model/make
3. Whether car is air-conditioned or not.
4. Condition of car.
5. Demand based on economic climate.

The likely impact of these factors on the selling price would be respectively:

1. Positive as CC rating increases.
2. Positive as reputation improves.
3. Air-conditioned prices would be higher.
4. Negative as condition gets worst.
5. As climate improves, demand for second-hand statsmobiles may decrease so prices may drop as well.

Judgement

O: At least six points (two or more factors + two or more impacts)
P: Between two and five points (factors plus impacts)

QUESTION FIVE

(a)

Evidence:
Between 1999 and 2009, the average annual average price of new five-door statsmobiles has:

1. Trended upwards then dropped in the last two years.
2. Peaked at $52.5K in 2007.
3. Ranged from $23.6K to $32.5K.
4. Risen on average by $3.6K annually over the years 1999 to 2007.
5. Decreased in the year 2001 from 2000 and in the year 2008 from 2007.
6. Increased on average by $2 100 per year overall.

Judgement

S: Three or more points correct.
P: One or two points correct.

(b)

Evidence:
In 2012, $x = 14$.
Using the line: Predicted y = $2.4745 \times 14 + 26.935 = \$61\ 600$
Using the curve: Predicted y = $-0.6186 \times 14^2 + 9.8983 \times 14 + 10.85 = \$28\ 200$

The line reflects the overall trend and gives an overestimate given the downwards trend since 2007. The curve reflects the most recent downward trend as well as the data in the first three years; however it doesn't allow for the seven years out of 10 where the price has been increasing.

So best estimate = average (line and curve) = $\frac{61.6 + 28.2}{2} = \$44\ 900$

Note:

1. For forecast look at average increase over 2002 to 2007 as \$1 000 per annum. So to 2012 use \$46.8K + \$3K = \$49 800.

Judgement:

S: Reasonable forecast with justification.
P: Reasonable forecast or some justification only.

(c)

Evidence:
If we calculate an index number series with 1999 = 1000 as the base year for 1999, 2003, 2007 and 2009 over the crude oil prices and annual prices for new five-door statsmobiles we obtain the following:

Year	Index Numbers		Comparison with 1999
	Statsmobiles	Crude Oil	
1999	1000	1000	
2003	2008	1518	2008 > 1518
2007	2225	4267	2225 < 4267
2009	1983	2983	1983 < 2983

Compared to 1999, over the years 1999 to 2003 the prices of new statsmobiles increased at a faster rate than oil prices. However over the years 2004 to 2009, the prices of oil increased at a faster rate than that of the new statsmobiles compared to 1999.

Note:

1. We could deflate the statsmobile prices to \$1 999. This gives rise to the deflated prices for the years 1999, 2003, 2007 and 2009 respectively as \$23 600, \$31, 600, \$12 300 and \$15 700. Compared to 1999, the prices of statsmobiles increased faster than oil prices in the period 1999 to 2003 however in the periods 1999 to 2007 and 1999 to 2009, the oil prices increased at a faster rate.
2. For the deflated price $15\ 700 = 46.8 \times \frac{21.49}{64.1}$

Judgement

O: Appropriate calculations and explanations.
P: One of either appropriate calculations or explanations.

Scoring for each Question

The codes are put together for each question and then converted to a mark out of eight according to the following table:

Mark	Codes
8	O + 2S, O + S + P
7	O, O + P, O + 2P, O + S
6	2S, 2S + P
5	S + P, S + 2P
4	S
3	3P
2	2P
1	P
0	N

The marks for each question are totalled to give an overall mark. Best possible overall mark is 40.